Janus Christian Jakobsen

Professor, Chief Physician

Copenhagen Trial Unit

Centre for Clinical Intervention Research

Rigshospitalet, Denmark

Quiz

- A trial is designed perfectly with predefined methodology per SPIRIT guideline
- The trial begins randomising patients

What are the greatest treaths to the validity of the trial results?

Quiz

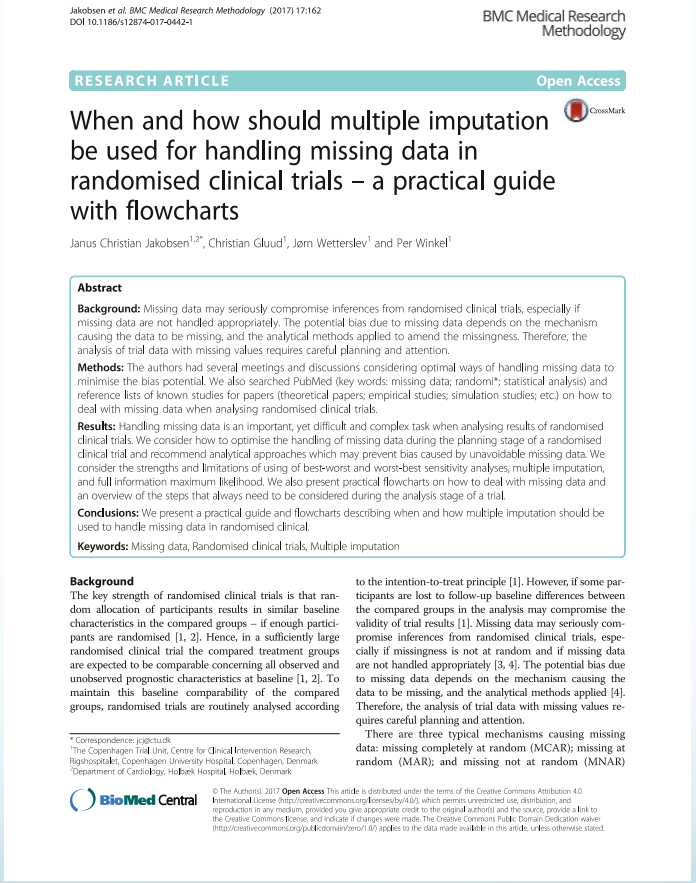
1. Not to reach the sample size
2. Missing data

Missing data in randomised trials

Types of missing data

- Patients do not receive the allocated trial intervention
  - NOT MISSING DATA !!
- Impossible to get the data, e.g., QOL assessment if a patient dies
- Withdrawel of informed consent (no use of data)
- Lost to follow-up

If missing data occur, can the problems be fixed ?



Jakobsen et al. BMC Medical Research Methodology (2017) 17:162
DOI 10.1186/s12874-017-0442-1

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts

Janus Christian Jakobsen[1,2*], Christian Gluud[1], Jørn Wetterslev[1] and Per Winkel[1]

Abstract

Background: Missing data may seriously compromise inferences from randomised clinical trials, especially if missing data are not handled appropriately. The potential bias due to missing data depends on the mechanism causing the data to be missing, and the analytical methods applied to amend the missingness. Therefore, the analysis of trial data with missing values requires careful planning and attention.

Methods: The authors had several meetings and discussions considering optimal ways of handling missing data to minimise the bias potential. We also searched PubMed (key words: missing data; randomi*; statistical analysis) and reference lists of known studies for papers (theoretical papers; empirical studies; simulation studies; etc.) on how to deal with missing data when analysing randomised clinical trials.

Results: Handling missing data is an important, yet difficult and complex task when analysing results of randomised clinical trials. We consider how to optimise the handling of missing data during the planning stage of a randomised clinical trial and recommend analytical approaches which may prevent bias caused by unavoidable missing data. We consider the strengths and limitations of using of best-worst and worst-best sensitivity analyses, multiple imputation, and full information maximum likelihood. We also present practical flowcharts on how to deal with missing data and an overview of the steps that always need to be considered during the analysis stage of a trial.

Conclusions: We present a practical guide and flowcharts describing when and how multiple imputation should be used to handle missing data in randomised clinical.

Keywords: Missing data, Randomised clinical trials, Multiple imputation

Background

The key strength of randomised clinical trials is that random allocation of participants results in similar baseline characteristics in the compared groups – if enough participants are randomised [1, 2]. Hence, in a sufficiently large randomised clinical trial the compared treatment groups are expected to be comparable concerning all observed and unobserved prognostic characteristics at baseline [1, 2]. To maintain this baseline comparability of the compared groups, randomised trials are routinely analysed according to the intention-to-treat principle [1]. However, if some participants are lost to follow-up baseline differences between the compared groups in the analysis may compromise the validity of trial results [1]. Missing data may seriously compromise inferences from randomised clinical trials, especially if missingness is not at random and if missing data are not handled appropriately [3, 4]. The potential bias due to missing data depends on the mechanism causing the data to be missing, and the analytical methods applied [4]. Therefore, the analysis of trial data with missing values requires careful planning and attention.

There are three typical mechanisms causing missing data: missing completely at random (MCAR); missing at random (MAR); and missing not at random (MNAR)

* Correspondence: jcj@ctu.dk
[1]The Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[2]Department of Cardiology, Holbæk Hospital, Holbæk, Denmark

Typical mechanisms causing missing data

- Missing completely at random (MCAR)
- Missing at random (MAR)
- Missing not at random (MNAR)

Multiple imputation

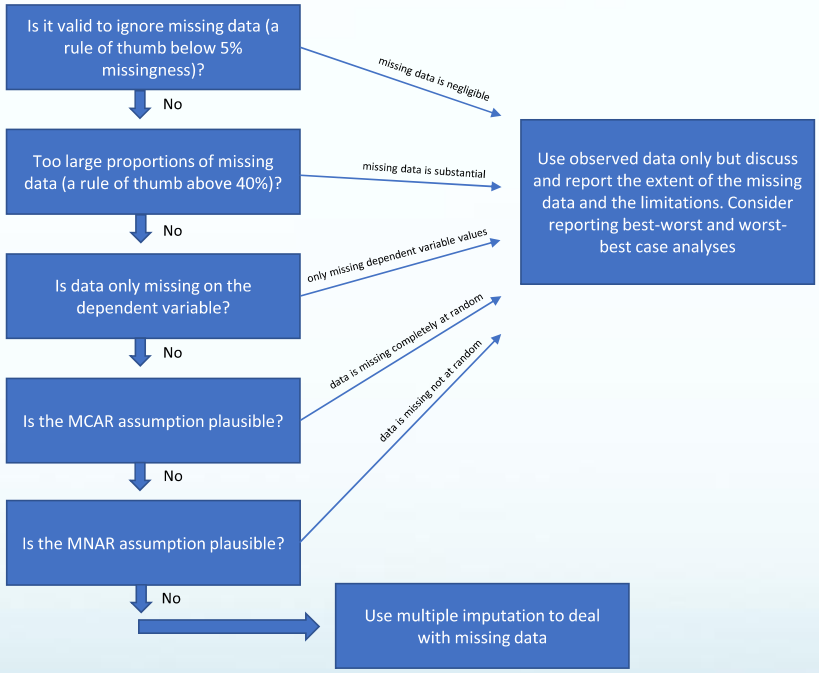


Fig. 1 Flowchart: when should multiple imputation be used to handle missing data when analysing results of randomised clinical trials

Best technique to handle missing data

- AVOID MISSING DATA
- Best-worst worst best analyses show the range of uncertainty

Missing data are one of the greatest threats to a well-designed trial

The biggest threat to the SafeBoosC-III follow up study?

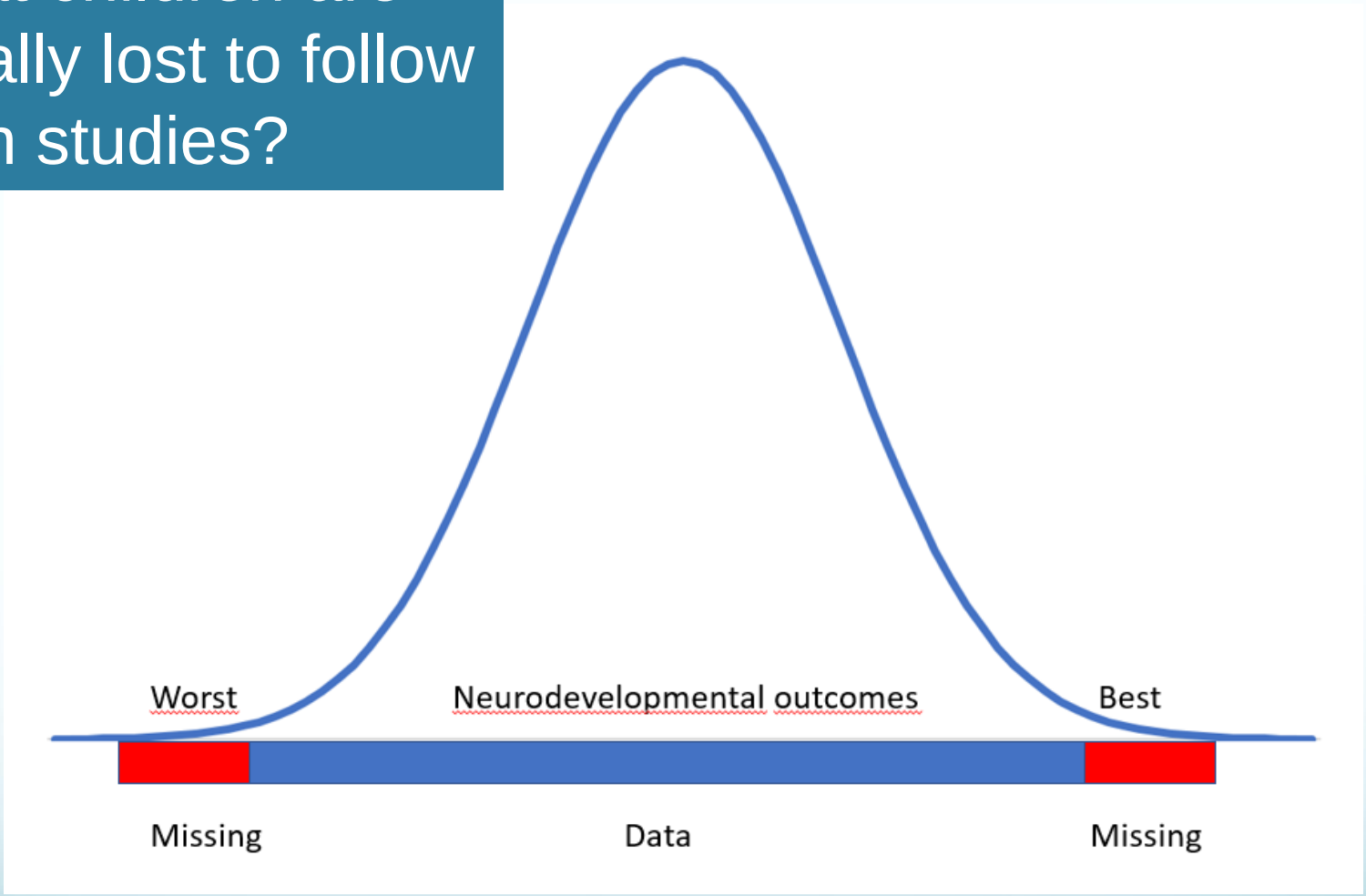
Missing data

Marie Isabel Skov Rasmussen, PhD student,
study manager of the SAfeBoosC-III follow up study

Loss to follow up in neonatal trials: what do we know?

- Families who are lost to follow up may have lower socioeconomic status
  - Maternal education
  - Maternal age
  - Foreign born
- Difficulties tracing families who move
- Lack of time due to other family obligations or work
- Financial barriers
- Not wanting to be reminded of the circumstances of the child's birth

What children are usually lost to follow up in studies?



Study (all published in NEJM)	Number of randomisations	Primary outcome	Follow up % at approx 2 years	Blinding
CAP trial: Long-term effects of caffeine therapy for Apnea of Prematurity	2006	Dichotomous: death or moderate to severe neurodevelopmental impairment at **18 months** based on clinical data	**93,2 %**	Not described
INIS trial: Treatment of Neonatal Sepsis with Intravenous Immune Globulin	3493	Dichotomous: death or major disability at **2 years of age** based on parental questionnaires	**96,7 %**	Double blinded
PENUT trial: A Randomised trial of erythropoietin for neuroprotection in preterm infants	941	Dichotomous: death or severe neurodevelopmental impairment at **22 to 26 months** based on clinical data	**79 %**	Double blinded
SIFT trial: Controlled Trial of Two Incremental Milk-Feeding Rates in Preterm Infants	2804	Dichotomous: survival without moderate or severe neurodevelopmental disability at **24 months** based on clinical data and parental questionnaire	**88 %**	Unblinded

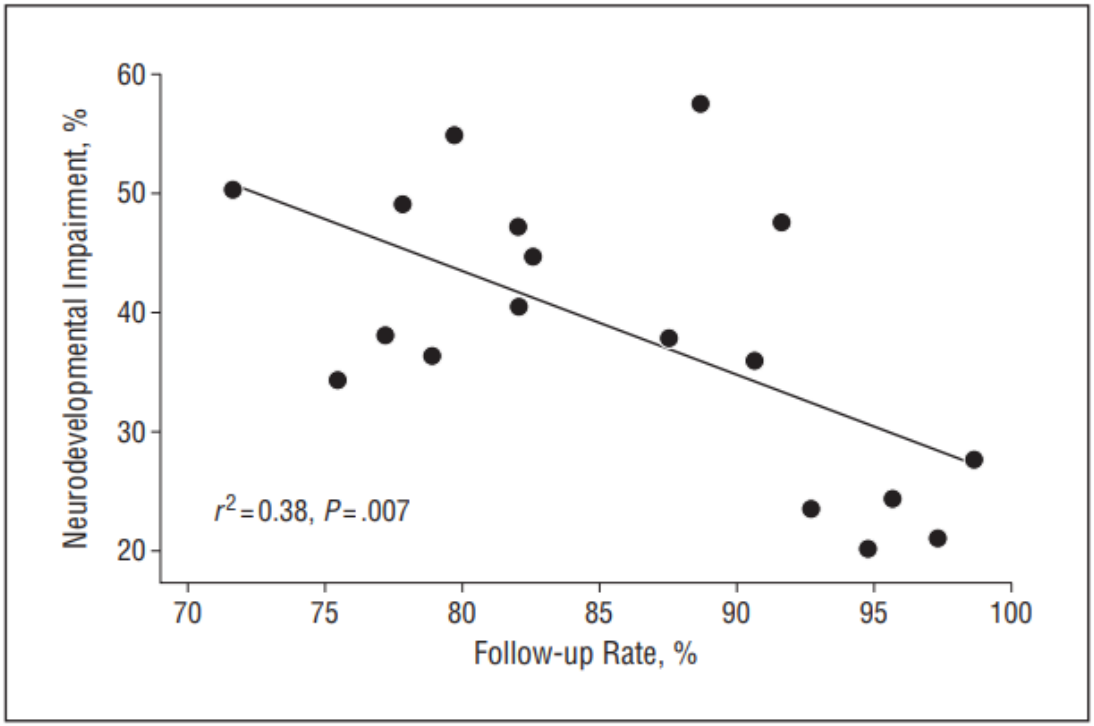


Figure 1. Scatterplot showing the relationship between neurodevelopmental impairment and follow-up rate. Solid line represents the linear regression fit across all subjects.

Guillén U, DeMauro S, Ma L et al. **Relationship between attrition and neurodevelopmental impairment rates in extremely preterm infants at 18 to 24 months: a systematic review.** Arch Pediatr Adolesc Med. 2012 Feb;166(2):178-84. doi: 10.1001/archpediatrics.2011.616.

Table 1 Ascertainment problems and incidence of severe disability at 2 years

Follow up	*Babies born in 1983*	*Babies born in 1990-1*	*Percentage with severe disability (95% confidence interval)*
Child reviewed without difficulty at 2 years	9/204	40/505	6.9 (5.2 to 9.0)
Child only traced with difficulty because of family mobility, etc	3/14	4/26	17.5 (7.3 to 32.8)
Assessment only arranged with difficulty after the child was traced*	6/12	20/35	55.3 (40.1 to 69.8)
Died after discharge from hospital but before the age of 2	1†/6	7†/16	36.4 (17.2 to 59.3)
Total	19/236	71/582	11.0 (8.8 to 13.1)

* Family mobility also added to the difficulty of maintaining contact with seven of these 47 families.
† Judged likely to become disabled by two years because of cerebral ultrasound and autopsy findings, and/or post-neonatal neurological symptoms.

Tin W, Fritz S, Wariyar U, Hey E. **Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty**. Arch Dis Child Fetal Neonatal Ed. 1998 Sep;79(2):F83-7. doi: 10.1136/fn.79.2.f83.

SIFT trial: 88% follow up (2470/2804)

Table S 5b: Sensitivity Analysis; Effect of missing data imputation on the primary outcome at 24 months CGA

Primary outcome at 24 months age corrected for prematurity	Faster arm (n = 1394)	Slower arm (n = 1399)	Unadjusted risk ratio (95% CI)	Adjusted risk ratio (95% CI)[a]	p-value[b]
Survival without moderate or severe disability[c], n/N (%)	802 / 1224 (65.5)	848 / 1246 (68.1)	0.96 (0.91, 1.02)	0.96 (0.92, 1.01)	0.160
Missing	170	153			
Assuming all infants with missing primary outcome do not have a moderate/severe disability, n/N (%)	972 / 1394 (69.7)	1001 / 1399 (71.6)	0.97 (0.93, 1.02)	0.97 (0.93, 1.02)	0.250
Assuming all infants with missing primary outcome do have a moderate/severe disability, n/N (%)	802 / 1394 (57.5)	848 / 1399 (60.6)	0.95 (0.89, 1.01)	0.95 (0.89, 1.01)	0.079
Assuming all infants in fast arm have a moderate/severe disability, while all infants in slow arm do not, n/N (%)	802 / 1394 (57.5)	1001 / 1399 (71.6)	0.80 (0.76, 0.85)	0.80 (0.75, 0.86)	<0.001
Assuming all infants in slow arm have a moderate/severe disability, while all infants in fast arm do not, n/N (%)	972 / 1394 (69.7)	848 / 1399 (60.6)	1.15 (1.09, 1.22)	1.15 (1.10, 1.21)	<0.001

[a]Adjusted for minimization factors; collaborating hospital, single or multiple birth, gestational age at birth, and birthweight less than the 10th centile for gestational age where technically possible.

[b]p-value for testing whether adjusted risk ratio is equal to 1.

[c]Moderate/severe disability is defined as one or more of the following: visual impairment, hearing impairment, motor impairment, or cognitive impairment (PARCA-R Composite Score <44). Definitions of motor and sensory impairments are defined in the report published by British Association of Perinatal Medicine (BAPM) in 2008.

Dorling J, Abbott J, Berrington J, and SIFT Investigators Group**. Controlled Trial of Two Incremental Milk-Feeding Rates in Preterm Infants**. N Engl J Med. 2019 Oct 10;381(15):1434-1443. doi: 10.1056/NEJMoa1816654.

How to minimise loss to follow up?

- Work hard to get both clinical data and parental questionnaires
- As soon as possible, think about what children might be in risk of loss to follow up
  - Identify the low ressource families and keep in contact with them
- Families who move - how do we get the data?
- Use the parental questionnaire
  - By phone?
  - At the follow-up clinic?